

University of Groningen

Recent developments in exponential random graph (p^*) models for social networks

Robins, Garry; Snijders, Tom; Wang, Peng; Handcock, Mark; Pattison, Philippa

Published in:
Social Networks

DOI:
[10.1016/j.socnet.2006.08.003](https://doi.org/10.1016/j.socnet.2006.08.003)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks*, 29(2), 192-215.
<https://doi.org/10.1016/j.socnet.2006.08.003>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Recent developments in exponential random graph (p^*) models for social networks

Garry Robins^{a,*}, Tom Snijders^b, Peng Wang^a,
Mark Handcock^c, Philippa Pattison^a

^a Department of Psychology, School of Behavioural Science, University of Melbourne, Vic., 3010, Australia

^b Faculty of Behavioral and Social Sciences, University of Groningen, The Netherlands

^c Department of Statistics, University of Washington, Seattle, USA

Abstract

This article reviews new specifications for exponential random graph models proposed by Snijders et al. [Snijders, T.A.B., Pattison, P., Robins, G.L., Handcock, M., 2006. New specifications for exponential random graph models. *Sociological Methodology*] and demonstrates their improvement over homogeneous Markov random graph models in fitting empirical network data. Not only do the new specifications show improvements in goodness of fit for various data sets, but they also help to avoid the problem of near-degeneracy that often afflicts the fitting of Markov random graph models in practice, particularly to network data exhibiting high levels of transitivity. The inclusion of a new higher order transitivity statistic allows estimation of parameters of exponential graph models for many (but not all) cases where it is impossible to estimate parameters of homogeneous Markov graph models. The new specifications were used to model a large number of classical small-scale network data sets and showed a dramatically better performance than Markov graph models. We also review three current programs for obtaining maximum likelihood estimates of model parameters and we compare these Monte Carlo maximum likelihood estimates with less accurate pseudo-likelihood estimates. Finally, we discuss whether homogeneous Markov random graph models may be superseded by the new specifications, and how additional elaborations may further improve model performance.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Exponential random graph models; p^* models; Statistical models for social networks

In recent years, there has been growing interest in exponential random graph models for social networks (Frank and Strauss, 1986; Frank, 1991; Wasserman and Pattison, 1996; see also Pattison and Wasserman, 1999; Robins et al., 1999). The class of exponential random graph

* Corresponding author.

E-mail address: garrylr@unimelb.edu.au (G. Robins).

models includes edge and dyadic independence models, the Markov random graphs of Frank and Strauss (1986), and many other graph distributions.

Markov random graphs are defined by a particular dependence structure between network ties, and have seemed a natural way to generalize the demonstrably limited dyad-independent models that preceded them. Although they have been the most widely explored form of these models to date, they have deficiencies that are now well-recognized (e.g. Snijders, 2002; Handcock, 2003b) and that are discussed further below. Models with higher order dependence assumptions have also been proposed (Pattison and Robins, 2002). In this article, we summarize new versions of exponential random graph models that generalize Markov random graphs. These new models incorporate the higher order specifications for exponential random graph models proposed by Snijders et al. (2006).

The motivation for the new specifications is the failure of Markov random graph models to fit observed social networks. These deficiencies had been masked previously by the use of approximate pseudo-likelihood estimation techniques. When we use more principled maximum likelihood estimation, it is often impossible to find satisfactory Markov graph parameter estimates (technically, the Markov Chain Monte Carlo (MCMC) estimation procedure does not *converge*). In these cases, Markov models cannot fit the data at all. The models are said to be *near degenerate*, a term we define and illustrate below. In this article, we show that, in terms of avoiding near degeneracy and the ability to obtain satisfactory parameter estimates, the new models perform dramatically better than Markov random graphs.¹

In summary, the results we describe below include the following: for 20 well-known data sets included with the UCINET program (Borgatti et al., 1999), Markov graph parameter estimates can be obtained for a little more than half, but estimation of the newly specified models is successful in every case. These UCINET networks are all relatively small. Our experience has indicated that it is increasingly difficult to obtain Markov model parameter estimates as observed networks get larger, whereas the newly specified models can still be estimated. An example of fitting the new specifications to a network with over a thousand nodes is given by Goodreau (this issue), a later paper in this special edition.

This is not to say that the new specifications provide satisfactory models in all cases. Rather, they are more successful than Markov models and offer the prospect of further ways forward to model complex networks. Moreover, just because the new specifications can be fit to given empirical network data, it is not necessarily the case that the resulting model is successful in reproducing all of the features of the observed network (although sometimes they do very well indeed). But with convergent parameter estimates for a particular model, a researcher is then in a position to add further effects to improve fit, as we illustrate below.

The structure of this article is as follows. We begin by presenting the general form of exponential random graph models. We then briefly describe Markov random graph models and the problem of near degeneracy to which they give rise. We then provide a simplified introduction to the new specifications of Snijders et al. (2006). We briefly outline estimation techniques and review currently available programs for maximum likelihood estimation. We then present the results of fitting both Markov and newly specified models to the UCINET networks, including a more detailed analysis for two of the data sets, and comparing maximum likelihood and pseudo-likelihood estimates for these cases. The paper concludes with remarks on further work.

¹ In what follows, we use the expressions, “non-convergent” and “near degenerate” (and, indeed, “degenerate”), somewhat interchangeably. The point is that, when a model is near degenerate, the parameter estimates will not converge, as described below.

1. Exponential random graph models

We use the notation and terminology described in Robins et al. (2007). For each pair i and j of a set N of n actors, Y_{ij} is a network tie variable with $Y_{ij} = 1$ if there is a network tie from i to j , and $Y_{ij} = 0$ otherwise. We specify y_{ij} as the observed value of Y_{ij} with \mathbf{Y} the matrix of all variables and \mathbf{y} the matrix of observed ties, the network. \mathbf{Y} may be *directed* or *non-directed*. A *configuration* is a set of nodes (usually small) and a subset of ties among them. For example, a *2-star* is a subset of three nodes in which one node is connected by a tie to each of the other two, and a *triangle* is a subset of three mutually connected nodes. Configurations are defined hierarchically, so that a triangle also includes three 2-stars.

The general form of the class of (homogeneous) exponential random graph models is as follows:

$$Pr(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} \exp\{\sum_A \eta_A g_A(\mathbf{y})\} \quad (1)$$

where:

- (i) the summation is over configuration types A ; different sets of configuration types represent different models (e.g. dyadic independence or Markov random graph);
- (ii) η_A is the parameter corresponding to configuration of type A ;
- (iii) $g_A(\mathbf{y})$ is the *network statistic* corresponding to configuration A (for homogeneous Markov graph models this is the number of configurations of type A observed in the network: for example, the number of triangles);
- (iv) κ is a normalizing quantity to ensure that (1) is a proper probability distribution.²

The model represents a probability distribution of graphs on a fixed node set, where the probability of observing a graph is dependent on the presence of the various configurations expressed by the model. One can interpret the structure of a typical graph in this distribution as the result of a cumulation of these particular local configurations. With suitable constraints on the number of parameters, it is possible to estimate parameters for a given observed network. The parameters then provide information about the presence of structural effects observed in the network data.

2. Markov random graphs

The *Markov random graphs* of Frank and Strauss (1986) are a particular sub-class of exponential random graph models in which a possible tie from i to j is assumed conditionally dependent³ only on other possible ties involving i and/or j . An example of a Markov random graph model for non-directed networks, with edge (or density), 2-star, 3-star and triangle parameters, is given in

² The above formulation of the model is specifically based on subgraph counts g_A because of a derivation from the Hammersley–Clifford theorem. This derivation is explicit in Frank and Strauss (1986) and is reflected in several recent treatments (e.g. Wasserman and Robins, 2005). An alternative approach is to assert various network statistics g as the basis for the model, whether they are subgraph counts or not. This was the approach taken by Wasserman and Pattison (1996) but has become less common. The advantage of the approach we follow here is that the dependence assumptions underpinning the model are thereby made explicit, so that the theoretical basis of the model is clear.

³ i.e. conditioning on the rest of the graph.

Eq. (2):

$$Pr(\mathbf{Y} = \mathbf{y}) = \frac{1}{\kappa} \exp\{\theta L(\mathbf{y}) + \sigma_2 S_2(\mathbf{y}) + \sigma_3 S_3(\mathbf{y}) + \tau T(\mathbf{y})\} \quad (2)$$

In Eq. (2), θ is the *density* or *edge* parameter and $L(\mathbf{y})$ refers to the number of edges in the graph \mathbf{y} ; σ_k and $S_k(\mathbf{y})$ refer to the parameter associated with k -star effects and the number of k -stars in \mathbf{y} ; while τ and $T(\mathbf{y})$ refer to the parameter for triangles and the number of triangles, respectively (Robins et al., 2007, describe this model and the edge, star and triangle effects). For a given observed network \mathbf{y} , parameter estimates indicate the strength of effects in the data. For instance, a large and positive estimate for τ suggests that, given the observed number of edges and stars, networks with more triangles are more likely; that is, there is a strong transitivity effect in the network. One of the strengths of these models is the explicit inclusion of transitivity effects, which are of course widely observed in social networks, but seldom successfully modeled (Snijders et al., 2006).

When the star and triangle effects are set to zero, so that the edge parameter θ is the only non-zero effect in the model, the result is a Bernoulli random graph distribution, often called the simple random graph or Erdős–Renyi graph distribution. For graphs in this distribution, edges occur independently of each other with a constant probability across the graph. Bernoulli random graphs are poor models for social networks, in part because they tend to have low levels of triangulation.

It is, in principle at least, relatively straightforward to simulate the distribution of graphs expressed in (2) for a given set of parameter values (and a fixed number of nodes), using for instance the Metropolis–Hastings algorithm. Strauss (1986) was the first to simulate Markov random graph distributions; more recent simulation results and additional descriptions of algorithms for simulation are in Snijders (2002), Handcock (2003a,b), Hunter and Handcock (2006) and Robins et al. (2005).

2.1. Near degeneracy

Simulation studies have revealed that for certain parameter values, models such as (2) have properties that lead us to question their value for data analysis. A graph distribution is termed *near degenerate* if it implies only a very few (possibly only one or two) distinct graphs with substantial non-zero probabilities (Handcock, 2003a,b). For certain parameter values, Markov random graph distributions exhibit this property, with only nearly empty or complete graphs likely under the distribution. These distributions cannot then be reasonable models for observed networks (which of course are usually neither empty nor complete).

Another problematic property may be a bimodal shape to the distribution, concentrated on two separate subsets of graphs, one of low density and one of density close to 1, with negligible probability for intermediate density graphs. This may be regarded as a softened version of near-degeneracy, the distribution being practically concentrated on two separated subsets rather than on only two outcomes. Unfortunately, degeneracy and bimodality are often observed when attempting to fit Markov models to networks where, for instance, transitivity is in the medium to high range, as is usual for social networks. This result calls into question whether Markov random graph models are plausible models for many empirically observed networks.

For more technical treatments, we refer readers to Handcock (2003a,b), Robins et al. (2005), Snijders (2002), Snijders et al. (2006) and Wasserman and Robins (2005). To illustrate the problem briefly, we present a simple example for graphs of 20 nodes (although the results generalize to

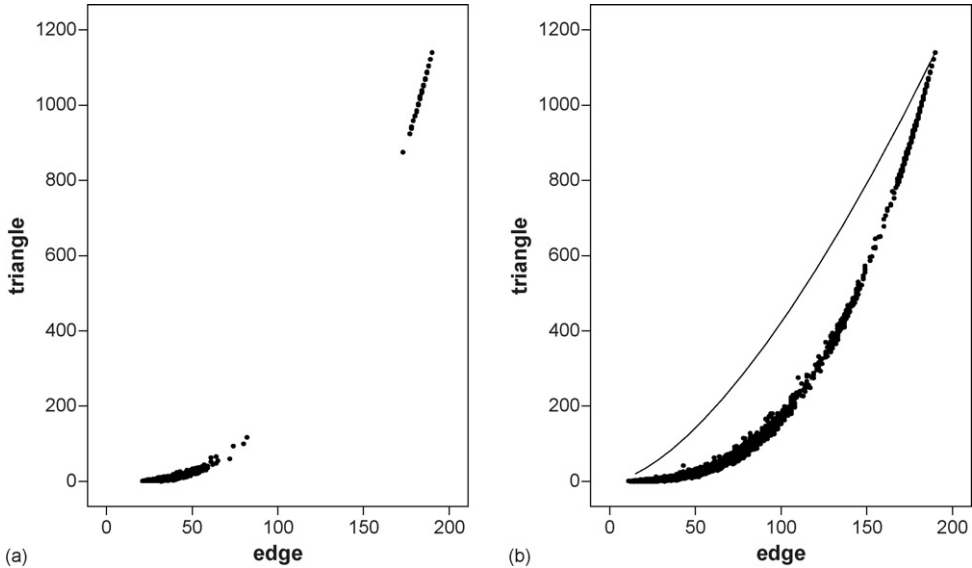


Fig. 1. Scatterplot for simulation study of edge/triangle Markov graph model: number of edges plotted against number of triangles for $\tau = 0.0$ – 1.0 in steps of 0.1 . Left panel (a) $\theta = -1.5$; right panel (b) $\theta = -2.0$ to 0.0 in steps of 0.5 .

any number of nodes). Fig. 1a plots the number of edges and triangles accumulated over multiple simulations of model (2). Specifically, a graph distribution is simulated for $\theta = -1.5$, $\sigma_2 = \sigma_3 = 0$, and for each value of the triangle parameter τ from 0.0 to 1.0 , in steps of 0.1 .⁴ The figure shows that the accumulated graph distributions contain a number of low density graphs corresponding to low values of τ (represented by the data points at the bottom left of the chart) and a well-separated number of high density graphs corresponding to higher values of τ (the data points at the upper right of the chart).

The graphs with few edges and triangles are in fact consistent with Bernoulli graph distributions, which at low density have very little triangulation. So, reassuringly, for low values of the triangle parameter we see few triangles in the graphs. The problem with this model arises, however, because a gradual increase in the triangle parameter does not lead to a gradual increase in graph triangulation. Rather we see a dramatic jump to very high-density graphs, graphs that are difficult to distinguish from high density Bernoulli graph distributions. In other words, changes to the triangle parameter have the effect of moving from a set of graphs that could have come from a (low density) Bernoulli graph distribution to a set of graphs that could have come from another (high density) Bernoulli graph distribution. Indeed, there is increased triangulation, but solely because the density of the graph has increased.

The key issue is that observed social networks tend to have somewhat low density but higher levels of triangulation than comparable Bernoulli graph distributions (see Snijders et al., 2006, for a discussion). In Fig. 1b, we have replicated the simulation of graph distributions described above, but this time allowing θ to range from -2.0 to 0 , in steps of 0.5 . This produces the curve of data

⁴ Each simulation involves 100,000 iterations and a graph is sampled at every 1000th iteration, resulting in a sample of 100 graphs. For each simulation, a burn-in of 50,000 iterations is implemented before the sampling begins. For each value of τ , two simulations are run, one with an empty starting graph and one with a full starting graph.

points at the bottom of the chart. This curve, however, is also difficult to distinguish with a range of Bernoulli graph distributions.⁵ The upper curve on the chart, on the other hand, indicates the maximum number of triangles in the graph possible for the given number of edges, so networks can have substantially more triangles than suggested by the lower curve of data points. (Some rather extreme combinations of θ and τ , when starting the simulations from a complete graph, have a non-negligible probability of leading to a simulated graph on the upper curve.) Because observed social networks so frequently have triangulation much greater than Bernoulli graphs, we expect most networks of interest to lie above the lower curve. Yet it is the lower curve that represents possible outcomes of the edge/triangle models we have simulated. This illustrates that a model with only edge and triangle effects is almost certainly inadequate for most observed social networks.

The problem may not be quite as bad as Fig. 1 suggests. For many observed networks, Markov graph distributions with non-zero 2- and 3-star parameters (and with the 3-star parameter negative) can produce reasonable models (see Robins et al., 2005; Snijders et al., 2006). Nevertheless, when networks have high concentrations of triangles (i.e. suggesting large positive τ), inclusion of star parameters is likely to be inadequate to prevent degeneracy or bimodality.

These problems do not apply only to the triangle parameter: there are degenerate regions of the parameter space also for the edge and 2-star model (Handcock, 2003b; Park and Newman, 2004).

3. New specifications

Snijders et al. (2006) proposed three new statistics that can be included in specifications for exponential random graph models: *alternating k-stars*, *alternating k-triangles* and *alternating independent two-paths*. For this article we concentrate on the first two. We particularly emphasize the utility of the new concept of alternating *k-triangles* as a higher order measure of transitivity. We typically refer to these parameters, and their associated models, as *higher order* because they include configurations with more than three nodes. In this section, we focus on higher order parameters for non-directed graphs, with some comments about directed versions at the end of the section.

3.1. Alternating *k-stars*

Technically, a model with an alternating *k*-star parameter alone remains within the class of Markov random graph models. The motivation for the new parameter comes from our experience of fitting models. In fitting models with star parameters to network data sets for which successfully convergent estimates could be obtained, we often observed that a higher order star parameter estimate was the opposite sign to, and smaller than, a lower order star parameter. For instance, it would be common in a Markov model such as (2) to see parameter estimates where σ_3 was approximately equal to $-\sigma_2/\lambda$ for some λ greater than 1.

In the Markov model of (2), the star parameters are deliberately limited to a maximum of 3-stars. More generally, however, the Markov dependence assumption permits higher order stars, up to stars of the maximum possible order ($n - 1$). In (2) there is an assumption that parameters

⁵ We know this because simulations with τ set to zero seem to cover the range of graphs produced by simulations with non-zero τ .

for these higher order stars are 0, the motivation being to limit the number of parameters and to achieve an identifiable model for which parameters can be estimated.⁶

The *alternating k -star assumption* proposes that, rather than setting higher order star parameters to equal 0, all star parameters be retained in the model but with a linear constraint among parameter values such that, for all $k \geq 2$, $\sigma_{(k+1)} = -\sigma_k/\lambda$ for some λ greater than 1. Then in (1) there is one parameter σ_2 that takes into account all star effects simultaneously, with an associated statistic:

$$u = \sum_{k=2}^{n-1} (-1)^k \frac{S_k}{\lambda^{k-2}} \quad (3)$$

In the context of the assumption that $\sigma_{(k+1)} = -\sigma_k/\lambda$ for $k \geq 2$, we refer to the parameter σ_2 as the *alternating k -star parameter*.

The statistic (3) includes stars of all orders but, for λ greater than 1, the impact of higher order stars is reduced for higher k (that is, for higher order stars) and they have alternating signs. In that sense, this assumption is rather more general than simply forcing higher order star parameters to have value 0, the assumption implicit in (2).⁷

An obvious question is what value to give to λ ? We have used $\lambda = 2$ in many cases, and this seems a reasonable starting point for many investigations. But of course it would be desirable to estimate an optimal value of λ . Hunter (2007), also in this special edition, shows how this can be done (see also Hunter and Handcock, 2006). For the purposes of this article, we set $\lambda = 2$.

What is important is how to interpret this parameter in terms of its implications for graph structure. Of course, the estimate of a parameter for a given local structure is usually indicative of the extent to which that structure is empirically observed (as long as there are no issues of degeneracy or bimodality). But the cumulation of these structures will produce particular global network patterns as well, so when a network exhibits such patterns, we might expect estimates of the relevant parameter to be large and significant.

If the alternating k -star parameter is positive, then highly probable networks are likely to contain some higher degree nodes (“hubs”), whereas a negative parameter suggests that networks with high degree nodes are improbable, so that nodes tend not to be hubs, with a smaller variance between the degrees. More particularly, Snijders et al. (2006) argued that a positive alternating k -star parameter (together with a negative density parameter) implied graphs that exhibit preference for connections between a larger number of low degree nodes and a smaller number of higher degree nodes, akin to a core–periphery structure. But once a node reaches a certain degree, the attainment of additional degrees adds little to its “popularity”.⁸ So for $\lambda = 2$, there is little difference between a node of degree 5, 6 or higher in “popularity”. Loosely, a node finds its way into the core once it has achieved a certain degree, with no particularly strong pressure for much higher degrees. At the same time, other nodes of lower degree remain outside the core. So the global implication is for a “loose” core–periphery structure, with few or any core nodes having particularly high degrees. This core–periphery structure is attained by (dampened) popularity processes. Changing the value

⁶ This assumption can be seen as analogous to ignoring higher order interactions in, for instance, a standard regression.

⁷ Note that in contrast to standard Markov graphs the statistic is no longer a direct count of configurations in the observed network, but rather a function of several configurations counts, in this case of counts of the various star configurations. In other words, we have utilized a Markov dependence assumption, but have imposed a particular form of constraint across star parameters, not the standard homogeneity constraint first used by Frank and Strauss (1986).

⁸ Here, what we are loosely terming as “popularity” of a node i can be considered more technically as the conditional log odds under the model of an additional node j forming a tie with i .

of λ changes the level of dampening, with higher λ implying increased chances for higher degree nodes. Later we show how core–periphery structures may also be attained by strong overlapping transitivity effects.

An important point is that the parameter assists in modeling the degree distribution. The Markov parameters in model (2) give some leverage over the first three moments of the degree distribution (via the edge, 2- and 3-star parameters), but these are not always sufficient when, for instance, degree distributions exhibit some outliers (high degree nodes). The new parameter, which does not exclude the higher order stars, provides leeway in dealing with such skewed degree distributions. This is not to say, of course, that all possibilities are accommodated, and when networks are highly centralized with a small number of very high degree nodes (hubs), a model with the new parameter alone is unlikely to provide a good fit to the data (we provide an example below). Nevertheless, our experience to date suggests that the alternating k -star parameter is more useful in model-fitting than a smaller number of individual Markov star parameters. Sometimes, as shown below, good fit of the degree distribution may be achieved with a combination of an alternating k -star parameter and a standard Markov 2-star parameter.

Why might the alternating k -stars assumption help with degeneracy? Snijders et al. (2006) argued that a model needs a balance between positive and negative star parameters to prevent the resulting graph distribution from being forced towards containing largely complete or largely empty graphs. The presence of alternating signs in the summation in (3) addresses this balance.

3.2. Alternating k -triangles

The alternating k -triangle assumption moves beyond the dependence assumptions underlying Markov random graph models, utilizing instead the partial dependence concept proposed by Pattison and Robins (2002). The underlying dependence assumption is presented by Snijders et al. (2006). In short, two possible edges in a graph, Y_{rs} and Y_{uv} , for distinct nodes, r, s, u, v , are assumed to be conditionally dependent if $Y_{su} = Y_{uv} = 1$. In other words, if the two possible edges in the graph were actually observed, they would create a 4-cycle. Substantively, it is not difficult to imagine situations where such *social circuit dependence* might be plausible. For instance, in an organization, suppose Steve typically worked with Ursula and Robyn with Victor. Then if Robyn and Steve commenced work on a new project, the chances of Ursula and Victor also working together might increase. In other words, the chances of Ursula and Victor working together are increased by Robyn and Steve working together, but only when Robyn already works with Victor and Steve with Ursula.

Snijders et al. (2006) invoke this dependence assumption *in addition to* Markov dependence, so in the above case, Y_{rs} and Y_{sv} (and so on) are also assumed to be dependent. It is not difficult to see intuitively that, with both social circuit and Markov dependence, configurations that are more complex than simple triangles are relevant to the model.

Snijders et al. (2006) proposed k -triangles as new higher order transitivity structures for non-directed graphs. A k -triangle is a combination of k individual triangles that all share one edge (the common *base* of the k triangles). Suppose that two adjacent nodes r and s are also both adjacent to the same k nodes t_1, t_2, \dots, t_k (i.e. there is an edge between r and t_m , and between s and t_m , for all $m = 1, 2, \dots, k$). Then the edge between r and s (the base) is common to the k individual triangles with nodes r, s and t_m . The entire structure comprising all the individual triangles is termed a k -triangle. The edges between r and t_m , and between s and t_m , are referred to as the *sides* of the k -triangle. Obviously a 1-triangle is an ordinary triangle. In Fig. 2, we depict 2-, 3- and 4-triangles.

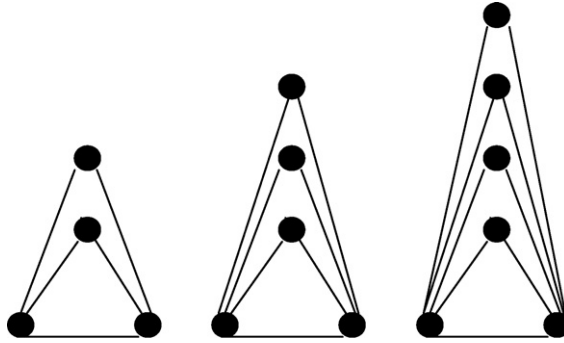


Fig. 2. 2-, 3- and 4-triangles.

As Snijders et al. (2006) showed, the inclusion of k -triangles in the model leads to a probability distribution where two edge variables are conditionally dependent only if either the Markov configuration (a common node) or the social circuit configuration (creation of a 4-cycle) applies. This is of some importance since it shows that the model remains strongly constrained by conditional independence assumptions.

Let T_k be the count of k -triangles in a graph. Then the k -triangles can be combined into one statistic in precisely the same way as for the alternating k -stars, that is:

$$v = \sum_{k=2}^{n-1} (-1)^k \frac{T_k}{\lambda^{k-2}} \quad (4)$$

This is the *alternating k -triangle statistic*. The corresponding *k -triangle parameter* $\tau = \tau_1$ corresponds to a (1-)triangle configuration, with the additional constraint that $\tau_{k+1} = -\tau_k/\lambda$, where τ_k is the parameter corresponding to a k -triangle. This statistic and its associated parameter are then included in models of the form (1).

The alternating k -triangle statistic does not simply represent triangulation in the network but additionally is a measure of the extent to which triangles themselves group together in larger higher order “clumps” in the network. Near degeneracy for Markov models often occurs when networks include larger clique-like structures that contain many triangles, rather than when many triangles are simply scattered rather homogeneously and individually throughout the network. In other words, networks that contain dense areas of clumping with many triangles are difficult to reproduce using for Markov random graphs. The detection and examination of such denser areas (e.g. the identification of *cohesive subsets of nodes*) has had an extensive history and remains a major theme in network analysis more generally. The transitivity parameter in Markov random graphs was a first attempt to model these denser structures because the triangle is the smallest clique among three nodes.

Yet, as Snijders et al. (2006) argued, a transitivity-based parameter incorporating subsets of nodes larger than three is needed to avoid degeneracy in modeling these agglomerations of triangles. Such parameters need to strike a balance between a tendency to have rather many cohesive subsets of three or more nodes, but a limitation of the total number of triangles, necessary to avoid a complete graph. As well, this parameter requires a statistic that is not linear in the triangle count inside the exponential, in contrast to (2). This linearity in the exponent is a major source of the degeneracy problem when graphs have many triangles. The alternating k -triangle statistic addresses these needs, with the alternating sign and decreasing weights for

higher order k -triangles serving the same purposes in limiting degeneracy problems as for alternating k -stars.

Broadly, a positive k -triangle parameter can be interpreted as evidence for transitivity effects in the network. More particularly, a positive k -triangle parameter suggests elements of a core–periphery structure (dependent on other effects in the model), but in this case due to triangulation effects rather than to popularity (degree) effects, as was the case for the alternating k -star effects. We might think of the star-based core–periphery structure as arising from the degree distribution, a natural outcome of a process whereby a subset of actors are more popular; whereas for the triangle-based core–periphery process, the degree distribution is itself one of the outcomes of the process, in which a core is built from overlapping triangulations. Below we also show how certain combinations of these star-based and triangle-based processes lead to network segmentation.

3.3. Alternating k -two-paths

Snijders et al. (2006) also proposed a parameter that is a lower order configuration for a k -triangle, namely a k -two-path. These configurations represent the number of distinct two-paths between a pair of nodes. They can be thought of as k -triangles without the base. The motivation is to introduce a parameter that, when used in conjunction with k -triangles, will enable researchers to distinguish between tendencies to form edges at the base, or at the sides of a k -triangle. Note that the sides of a k -triangle in the absence of the base represent a type of edge clustering that is only the precondition to transitivity, while the presence of the base edge reflects transitive closure. So this combination of parameters can provide evidence for pressures to transitive closure (but pressures that do not increasing linearly with k in the exponent). Fig. 3 depicts 2-, 3- and 4-two path structures. Counts of these structures are then incorporated into one statistic analogously to (4). In this article, we do not concentrate on alternating k -two-path parameters. For some data, we have found it important to include them in models but further work is needed to understand better their effect when included with other parameters.

3.4. Alternative equivalent parameterizations

There are equivalent parameterizations to the alternating k -star, alternating k -triangle and alternating k -two-path specifications for the model that have mathematical elegance and enhance interpretation.

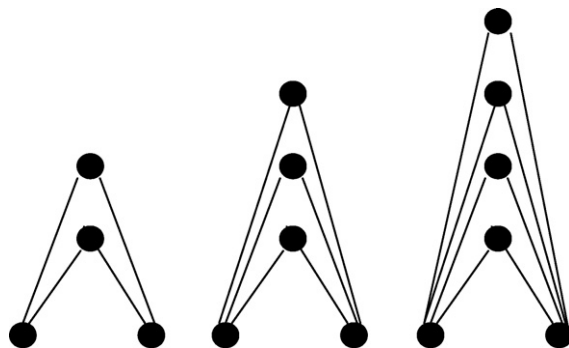


Fig. 3. 2-, 3- and 4-independent two-paths.

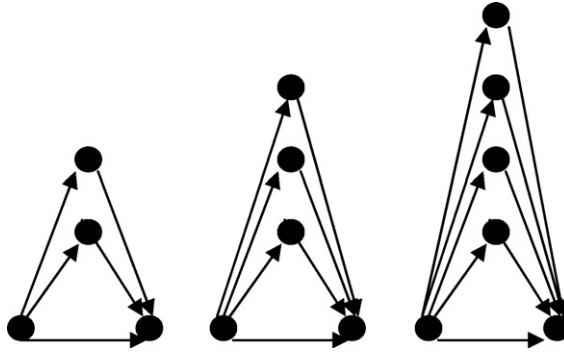


Fig. 4. Directed 2-, 3- and 4-triangles.

1. The *geometrically weighted degree parameter* explicitly models the degree distribution but puts more weight on the numbers of nodes with lower degrees, with weights decreasing geometrically as the degrees increase.
2. The *edge-wise shared partner* (or ESP) parameter models the distribution of shared partners of tied actors, but with weights decreasing geometrically as the number of shared partners increase.
3. The *dyad-wise shared partner* (or DSP) parameter models the distribution of shared partners of actors who may or may not be tied, but with weights decreasing geometrically as the number of shared partners increase.

These parameterizations are presented in detail in [Hunter \(2007\)](#), who explains the connections between the two equivalent parameterizations, so we do not discuss them further here.

3.5. New specifications for directed graphs

The above descriptions apply to non-directed graphs. [Snijders et al. \(2006\)](#) also proposed counterparts for directed graphs. They suggested two versions of the alternating k -star parameter, one based on in-stars and one on out-stars, referred to as *alternating k -in-stars* and *alternating k -out-stars*, respectively. In all other respects, these are directly analogous to the non-directed versions. For directed k -triangles, they proposed a structure in which the base directly connects two nodes that are otherwise reachable by k directed two-paths. [Fig. 4](#) depicts directed 2-, 3- and 4-triangles. These are directed k -triangles as referred to in this article. Of course, there are other possibilities where the two nodes at the base are otherwise reachable by different types of two-semi-paths. We will not discuss these possibilities here, but work on these various types of directed k -triangles is in progress.

4. Estimation

Estimation techniques for these models have been recently discussed by various authors (e.g. [Hunter and Handcock, 2006](#); [Snijders, 2002](#); [Wasserman and Robins, 2005](#); see [Robins et al., 2007](#), for a summary), so we will restrict our comments here to summary remarks. To date, the most common form of estimation for Markov random graph models has been maximum pseudo-

likelihood (Strauss and Ikeda, 1990). The properties of the pseudo-likelihood estimator are not well understood, the pseudo-likelihood estimates can at best be thought of as approximate, and it is not clear from existing research as to when pseudo-likelihood estimates may be acceptable. We shall shed some light on this for actual data below.

Monte Carlo Markov chain maximum (MCMC) likelihood estimation, when available, is the preferred estimation procedure.⁹ The different Monte Carlo techniques of Snijders (2002) and Hunter and Handcock (2006) are both based on refining approximate parameter estimates by comparing the observed graphs against a distribution of random graphs generated by a stochastic simulation using the approximate parameter values. If the parameter estimates never stabilize (*converge*), the model is likely to be degenerate. When convergent estimates are obtained, then simulation from the estimates will produce distributions of graphs in which the observed graph is typical for all of the effects in the model. One of the advantages over maximum pseudo-likelihood estimates is that one can also obtain reliable standard errors for the estimates.

Snijders et al. (2006) conditioned on the number of edges when estimating parameters, that is, they fixed the number of edges in Monte Carlo estimation procedures (see Snijders and van Duijn, 2002; also Frank and Strauss, 1986, who proposed fixing parameters for estimation of Markov graphs). In such models there are no density parameters. Fixing the number of edges is designed to diminish the risk of degeneracy problems and will have minor effects on other parameter estimates (except perhaps for star parameters). Our experience has been that, at least with smaller networks, conditioning on edges may not be necessary, and estimation procedures may successfully converge for the new specifications with density parameters included. This is the case for all of the examples of convergence that we present in this article.¹⁰

4.1. Programs for Monte Carlo maximum likelihood estimation

There are currently three programs available for Monte Carlo maximum likelihood estimation. SIENA and pnet were used to obtain the results in this article. Both use the same algorithm as described below. The program *statnet* is used in other articles in this special edition (Goodreau, 2007; Hunter, 2007).

4.1.1. SIENA (and pnet)

The first generally available software to implement Monte Carlo maximum likelihood estimation for exponential random graph models was SIENA (in its 2002 release; the current version is Snijders et al., 2005) in the StOCNET suite of programs (Boer et al., 2006). The SIENA procedure implements the stochastic approximation algorithm described in Snijders (2002). More recently, the pnet program (Wang et al., 2006) has also been developed, implementing the same algorithm. The estimates from the two programs have been checked against each other, ensuring they produce the same results (except for differences caused by the stochastic nature of the algorithm). In addition to estimates and standard errors, output includes a convergence *t*-ratio for each estimate. This ratio indicates how well that estimate has converged, with a small value close to

⁹ At the time of writing, Monte Carlo estimation was not available, for instance, for some of the more complicated exponential random graph models in the literature, e.g. for valued networks, for multiple network models involving more than two networks, and for tripartite graphs. Work is progressing in developing the programs to estimate more complicated models.

¹⁰ But this may not be universally so. For larger networks, especially directed networks, in some instances conditioning on edges may be required for convergence. Arguably, this is an indication that model specification is not exactly right.

zero suggesting good convergence. Estimation may be conditional on edges, or unconditional. If any of the t -ratios are too large, this may indicate either that convergence is not yet satisfactory or that the model is degenerate and incapable of fitting the data. For non-degenerate models, good convergence is indicated by t -ratios for all parameter estimates being less than (or close to) 0.1 in absolute value. (The convergence t -ratio should be distinguished from the usual t -statistic defined as estimate divided by standard error, and used for testing whether the coefficient is 0.)

Efficiency has been improved in the *pnet* program to the point where on a standard laptop computer convergent estimates for models with higher order parameters can be obtained in minutes for networks of, for instance, 150 nodes. For very large graphs of one thousand or more nodes, estimation will be slow (and *statnet* might be the preferred option).

Both SIENA and *pnet* contain some diagnostics that in addition to the convergence t -ratios will assist judgments about convergence. For instance, SIENA includes the hysteresis analysis described in Snijders et al. (2006). Another good diagnostic is to simulate from the parameter estimates to check that the observed graph is not extreme in the model statistics derived from the simulated distribution. Sometimes the simulation shows two regions of possible graphs (e.g. high density and low density as in Fig. 1a), indicating degeneracy.

Additional information is available from the SIENA website (<http://stat.gamma.rug.nl/siena.html>) and the StOCNET homepage (<http://stat.gamma.rug.nl/stocnet/>). Material on *pnet* is available from the Melnet website (<http://www.sna.unimelb.edu.au/>).

Statnet is an integrated set of software tools for the representation, visualization, simulation and analysis of network data. The package allows maximum likelihood estimates of exponential random graph models to be calculated using Markov Chain Monte Carlo. The primary method is MCMC Newton–Raphson as described in Geyer and Thompson (1992). The *statnet* algorithm is different from the stochastic approximation algorithm of Snijders (2002) and is intended to optimize computational efficiency in estimation. For instance, rather than having many simulation runs within the program, as is the case for SIENA and *pnet*, it implements very few simulation runs (possibly only one) but then weights the graphs within the simulations in ways that optimizes estimation (see Handcock, 2003b; Hunter and Handcock, 2006, for details). This means that the program is able to estimate models for large graphs in a reasonably short time frame. The computational procedures in *statnet* are sophisticated, incorporating recent advances in Metropolis–Hastings algorithms and multiple proposal types that provide flexibility for fitting many kinds of models. The software is written for computing environment *R*, a GNU-licensed freeware statistical computing package that runs on a variety of platforms, including Windows, Linux and Macintosh based systems. This program uses the *R* statistical language as an interface and so enables a researcher to employ the extensive features of *R* in an analysis. If, on the other hand, a researcher wishes simply to estimate models, the *R* commands to do this are not difficult to learn. The *statnet* package does much more than provide parameter estimates. It provides a framework for the analysis of network data. It uses the *network* package to store network data (Butts, 2005). The bundle also provides simple tools for plotting networks, simulating networks and assessing model goodness of fit (see Goodreau, 2007). The performance of the MCMC algorithms can be monitored graphically and numerically. The summary of model results includes the usual estimates and standard errors, as well as estimates of the error attributable to the MCMC algorithm and an analysis of the statistical deviance (akin to GLM models). In addition, many of the traditional methods and tools of social network analysis are available via the *sna* package (Butts, 2002). Additional material is on the *statnet* webpage <http://csde.washington.edu/statnet> and in Handcock et al. (2004).

5. Models with combinations of parameters

There is no impediment to using Markov parameters and the newly specified parameters together in one model (we give an example below) although the precise interpretation of any parameter depends on the other effects in the model. For this article, we wish to compare the performance of each, so for the most part we keep Markov and the new parameters in separate models.

As with any modeling endeavor, the application of one model to the data may not be the end of the story. Often it is desirable to fit several different models, either to explore different parameter combinations if it happens that convergence is difficult to achieve, or to find the model that best represents the data. For the purposes of this article, for non-directed graphs we included 2- and 3-star and triangle parameters in Markov models, and alternating k -stars and alternating k -triangles in higher order models. If these models did not converge we tried either simpler or more complex models (e.g. with additional or fewer star parameters). We often found, for instance, that for non-directed networks very simple models with only edge and alternating k -triangles parameters successfully converged. For Markov models for directed networks, we began by including density, reciprocity, 2- and 3-in- and out-star parameters, and 2-mixed-star, and triangle parameters; and for higher order models for directed networks, we included density, reciprocity, alternating k -in-stars, alternating k -out-stars and directed alternating k -triangles.

Many of these statistics may exhibit quite high levels of collinearity. Indeed the alternating k -star and alternating k -triangle statistics can be shown to include the number of edges as one of the terms in a closed form expression for the statistic (Snijders et al., 2006; see also Hunter, 2007). As the alternative weighted degree parameterization does not include the number of edges (Hunter, 2007), this might be a reason to prefer it. In practice, however, collinearity is not usually a problem for estimating these models, as long as appropriate Monte Carlo estimation techniques are used.

5.1. Model interpretation and “goodness of fit”

We have already discussed some aspects of parameter interpretation. In brief, a positive alternating k -star parameter estimate suggests a degree distribution containing some higher degree nodes, and a resulting “loose” core–periphery structure; whereas a negative parameter suggests a truncated degree distribution with a tendency against particularly high degree nodes. A positive k -triangle parameter estimate suggests a tendency for triangulation in the graph, with the triangle tending to form together into “clumps”.

For the higher order models, it is important to interpret the combination in the one model of alternating k -stars and alternating k -triangle parameters. A pattern of negative k -star and positive k -triangle estimates is not uncommon. Here there are two countervailing tendencies: one towards a triangulated core–periphery structure, and one against a degree-based core–periphery structure. The global outcome is not a single core of one internally densely connected set of nodes, but several (often connected) smaller regions of overlapping triangles. An example is provided in Fig. 5. A single core–periphery structure implied by the positive k -triangle effect has been segmented by the inclusion of negative k -star parameter into a chain of smaller dense regions of the network.¹¹

¹¹ This graph is from a simulation of non-directed graphs on 65 nodes with parameter values: edge = -0.5 , alternating k -star = -1.5 , alternating k -triangle = 2.0 .

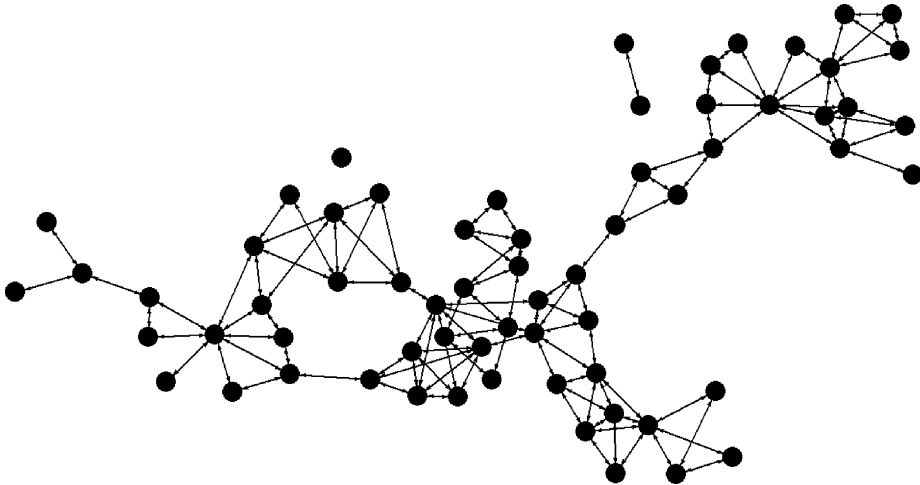


Fig. 5. A network based on a positive alternating k -triangle effect and a negative alternating k -star effect.

So a range of models with a fixed positive k -triangle parameter, but with k -star parameters ranging from 0 to increasingly negative values, sees a move from centralization to segmentation in the network.

Once parameters have been estimated from an observed network, it is often useful to simulate the resulting distribution of graphs, not only to check whether the observed statistics are well reproduced by the model, but also to investigate how well the model parameters succeed in replicating features of the observed graph that are not explicitly modeled. This process can be construed as a “goodness of fit” investigation, and may suggest refinements to the model, for example, the inclusion of additional parameters.

Given that a good model should be able to reproduce the data, it can be useful to consider those aspects of the observed graph that are not explicitly modeled. For instance, if a model includes a reciprocity parameter, a good set of estimates will naturally reproduce the number of reciprocated ties in the graph with a high degree of accuracy. But, irrespective of the model, the graph will still have many features that are not explicitly, or only partially, modeled. For instance, the degree distribution is only partially modeled by k -stars, the distribution of shortest path lengths between pair of nodes (that is, the geodesic distribution) is not explicitly modeled at all, and so on. If a simulation based on the model estimates can reproduce these additional features (to some reasonable extent), then our confidence increases that this model is indeed an effective representation of the data. Moreover, we could argue that these additional features may have emerged from the processes that are implicit in the model parameterization, that is, there is no evidence that other effects are necessary to explain these features of the observed network.

To compare a simulation of the model against non-parameterized features of the graph is in fact a rather difficult test of how well the model “fits” the data, and it goes a considerable way beyond what is usual in more standard statistical approaches. But it is our experience to date that these models can often reproduce a range of features of the observed graph surprisingly well, at least in the sense that observed statistics are not extreme compared with distributions generated by the simulation.

This is not just an issue of model fit. A simulation may also enhance interpretation of the model. If we simulate from parameter estimates and produce a sample of graphs that typically

have a long tailed degree distribution, or a high clustering coefficient, we can infer that this is an expected outcome of the model. Goodreau (2007) provides an example of the effectiveness of post hoc simulations in assessing and interpreting a fitted model.

6. Fitting the new specifications to UCINET data sets

We fitted Markov models and models with the new specifications to 20 well-known data sets available with the UCINET V network analysis program (Borgatti et al., 1999). The networks we examined were one-mode (not bipartite), binary (not rankings) and positive relationships (not antagonistic). The data sets comprised: the Kapferer mine data (kapfmm and kapfmu, both of 16 nodes and non-directed); the Kapferer tailor shop data (all of 39 nodes, with two non-directed networks – kapfts1 and kapfts2 – and two directed networks – kapfti1 and kapfti2); the Wolf primate networks (one directed network, wolfk, on 20 nodes); the Padgett Florentine families data (two non-directed networks on 16 nodes, padgm and padgb); the Krackhardt hi-tech managers data (two directed networks on 21 nodes, friendship and advice); the Read Highland tribes network (one non-directed network on 16 nodes, Gamapos); the Zachary karate club network (one non-directed network on 34 nodes, Zache); the Roethlisberger and Dickson bank wiring room data (all on 14 nodes, with two non-directed networks – rdpos and rdgam – and one directed network, rdhlp); the Taro exchange network (one non-directed network on 22 nodes); the Thurman office network (one non-directed network on 15 nodes); and the Knoke bureaucracies data (two directed networks on 10 nodes, knokm and knoki).¹² In summary, we have 12 non-directed networks and 8 directed networks, with numbers of nodes ranging from 10 to 39.

Models were fitted to all data sets using pnet and checked with SIENA. Without going through the details of all the results, Markov models could not be estimated for 5 of 12 non-directed networks (kapfts1, kapfts2, gamapos, zache, taro), and for 3 of 8 directed networks (kapfti1, kapfti2, wolfk). Models incorporating the new specifications, on the other hand, were estimable for all of the data sets. All of these higher order models included parameters for k -triangles, with the exception of the wolfk network which included parameters for alternating in- and out-stars but was degenerate when an alternating k -triangle parameter was included.¹³

6.1. Some particular examples: estimates and goodness of fit

Rather than including details of parameter estimates of all models, we provide two illustrative examples: one for a non-directed network where Markov models converge (padgb), and one where Markov models are degenerate (kapfts1) (Table 1).

For the Padgett Florentine families business network (padgb), parameter estimates, standard errors and convergence t -ratios for both Markov and higher order models are presented in Table 2. In contrast to pseudo-likelihood estimation, these standard errors are meaningful, so a parameter estimate that is more than twice its standard error can be considered significantly different from

¹² For more detailed descriptions of the data, together with references to the original works, see Borgatti et al. (1999).

¹³ It is perhaps not surprising that this network represents special difficulties for any model with triangle parameters. The network represents kinship relations among a group of primates, and contains only one triangle, as well as no mutual ties. As such it is a rather unusual “social” network. In Siena, it is possible to keep the dyad count (the counts of mutual, asymmetric and null dyads) fixed in the estimation. This is appropriate for the wolfk network because it restricts the model to networks without mutual ties. When this is done, estimation for the wolfk network converges both with the Markov and with the new specifications.

Table 1
Models for the padgb network: parameter estimates, standard errors and convergence statistics

Parameter	Estimate	Standard error	Convergence-Statistic
Markov model			
Edge	−4.273	1.127	0.03
2-star	1.093	0.650	0.01
3-star	−0.673	0.406	−0.02
Triangle	1.316	0.646	−0.02
Higher order model			
Edge	−2.754	0.854	0.01
Alternating <i>k</i> -stars	−0.064	0.485	−0.02
Alternating <i>k</i> -triangles	0.856	0.468	−0.06

zero. For the triangle and alternating *k*-triangle parameters, a one-sided test is appropriate in most applications, so estimates exceeding 1.65 times the standard error can be regarded as significant at the 5% level.

In each model, the edge (or density) parameter is significant, as is the triangle parameter in the Markov model and the alternating *k*-triangle parameter in the higher order model. It would be particularly naïve simply to exclude non-significant parameters from the model without a little forethought. If we did so, for instance, in the Markov model, we would revert to the edge/triangle model which is never likely to be an adequate representation of the data (see Fig. 1).

Our interpretation of the parameters is as follows. For the Markov model, the large negative edge parameter indicates a low baseline propensity (net of other effects) to form ties. The positive 2-star and negative 3-star parameters in combination indicates a tendency for multiple network partners but with a ceiling on this effect, so that while most nodes will have 2, 3, or possibly more partners, and the variance of the degrees may be higher than expected for a random graph, the number of very high degree nodes is limited. The positive triangle parameter suggests tendencies for transitivity in this network. We can see these effects in the network as depicted in Fig. 6. There

Table 2
Models for the padgb network: goodness of fit statistics

Effect	Model		
	Bernoulli	Markov	Higher order
Graph counts			
Edges	0.15	0.03	−0.05
2-stars	0.95	0.03	−0.13
3-stars	1.04	0.03	−0.29
Triangles	3.16	0.00	−0.13
Alternating <i>k</i> -stars	0.83	0.02	−0.03
Alternating <i>k</i> -triangles	3.06	0.23	−0.01
Degree distribution			
Standard deviation	2.29	0.54	0.39
Skew	−0.05	0.08	−0.28
Clustering			
Global	2.92	0.24	0.32
Local	3.52	0.93	0.66

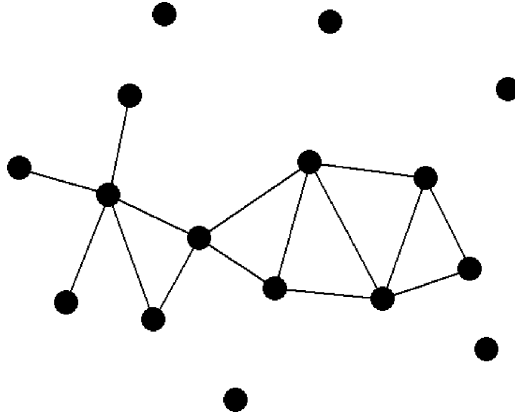


Fig. 6. The padgb network.

is an obviously high level of triangulation, and a tendency for non-isolated nodes to have multiple partners, but the maximum number of partners is no more than five.

In the higher order model, the positive alternating k -triangle parameter also suggests transitivity. Additionally, it indicates that triangles tend to be overlapping in denser regions of the network. The fact that the alternating k -star statistic is not large has two implications: first, there seems to be some level of a core–periphery tendency here, and not for a more segmented network; and, secondly, the core–periphery tendency is due to triangulation, not popularity effects. We can also infer that once the core–periphery triangulation is accounted for, star effects (present in the Markov model) are not necessary to explain the data.

Table 2 presents goodness of fit statistics for various models for this network. The statistics relate to several features of the graph, not just those statistics expressed by the models. The additional features include standard deviation and skewness of the degree distribution, and local and global clustering coefficient.¹⁴ The goodness of fit (‘gof’) statistics are calculated from the means and standard deviations for the relevant measures from a sample of graphs simulated from the parameter estimates. Each gof statistic is the difference between the measures for the observed graph and the simulated mean, scaled by the standard deviation.¹⁵ A small gof statistic indicates that the model explains this particular feature of the network well. In addition to Markov and higher order models, for sake of completeness, Table 2 also includes goodness of fit statistics for a Bernoulli model (i.e. edge parameter only).

We see from Table 2 that the Bernoulli model is not a good fit to the degree distribution and any feature that relates to clustering including triangle and k -triangle counts. However, both Markov

¹⁴ The global clustering coefficient is three times the number of triangles in the graph as a ratio of the number of two-stars. The local clustering coefficient is calculated as follows. For each node i the local density is the density among all nodes adjacent to i . The local clustering coefficient is then the mean of these local densities across all i .

¹⁵ These are identical calculations to a t -statistic, based on the means and standard deviations from the simulated graph distribution. Thus, they are formally identical to the convergence t -ratios but fulfill a different role: they assess the goodness of fit for aspects of the network not explicitly included in the model, rather than checking convergence of the estimation algorithm. For the fitted parameters, any difference between the t -ratio used for convergence, and the gof statistic, is due to the stochastic nature of the Monte Carlo estimation approach. We infer that the gof statistics are not extreme if they are less than 2 in absolute value, but we view this as a heuristic approach to model examination, rather than a formal hypothesis test (for one thing, hypothesis tests for the one set of parameters would not all be independent).

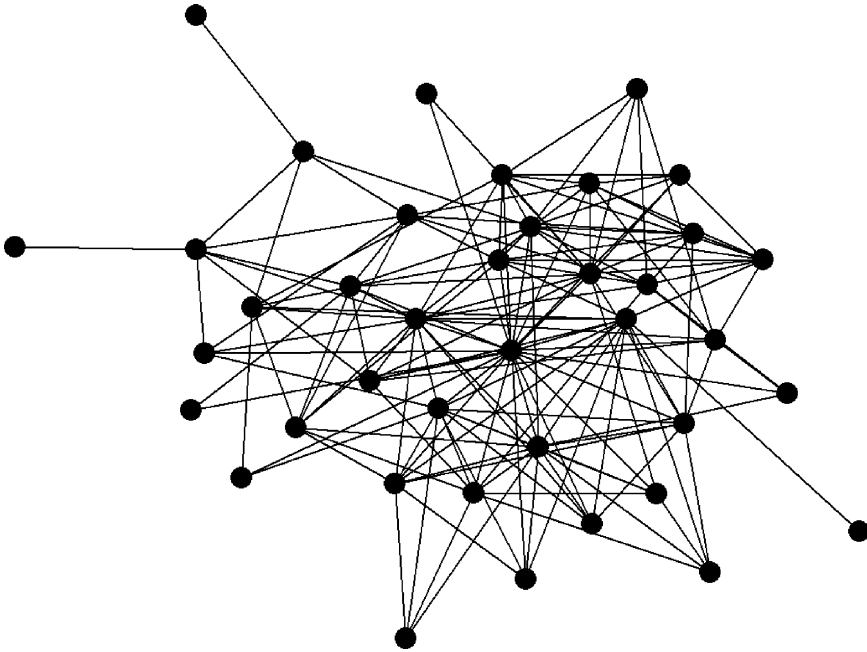


Fig. 7. The kapfts1 network.

and higher order models seem to fit well on all of the features. In addition, we have examined the quartiles of geodesic distributions: both Markov and higher order models produce graphs with geodesic quartiles that are consistent with the observed network. This is a case where a Markov model seems quite adequate and where the higher order model does not provide any marked gains in terms of fitting the data, although as discussed above consideration of both models provides some interesting points on interpretation.

Next, we turn to the example of the Kapferer tailor shop data, *kapfts1*, for which the estimation procedure did not converge in the Markov model case. The network is depicted in Fig. 7. The first point that becomes clear is that because of the larger number of nodes compared with the previous example, it is not simple to make inferences based solely on visual inspection.

The top panel of Table 3 provides parameter estimates for the higher order model for this data: goodness of fit statistics are provided in the left column of the third panel. In this example we have included the k -two-path parameter in the models but as can be seen this effect is not substantial.

This simple four-parameter model indicates significant clustering among the nodes into dense regions of overlapping triangles. The fact that the alternating k -star parameter, although negative, is small and not significant, suggests a triangulated core–periphery structure. These features are apparent from Fig. 7. But what is also clear from the goodness of fit is that this model does not do a particularly good job of explaining the degree distribution. The observed graph has a more dispersed and skewed degree distribution than the model would suggest. As well, the level of clustering in the observed graph is even higher than that predicted by the model. So, for this data, while we can determine nothing from a Markov model, we do learn something from the higher order specifications, but there are additional features of the network that are not captured.

In some cases it is possible to incorporate further additional effects in the model to improve fit (e.g. Snijders et al., 2006). For the *kapfts1* network, given the failure to provide a good fit to the

Table 3

Models for the kapfts1 network: parameter estimates, standard errors, convergence and goodness of fit statistics

Parameter	Estimate	Standard error	Convergence-Statistic
Higher order model			
Edge	−3.779	1.495	−0.04
Alternating k -stars	−0.081	0.409	−0.04
Alternating k -triangles	1.129	0.206	−0.04
Alternating k -two-paths	0.029	0.054	−0.05
Higher order model with 2-star parameter			
Edge	−1.961	1.451	0.00
2-star	0.046	0.017	−0.01
Alternating k -stars	−0.589	0.412	0.00
Alternating k -triangles	0.952	0.216	0.00
Alternating k -two-paths	−0.007	0.034	0.01
	H/O model		H/O model with 2-stars
Goodness of fit statistics			
Graph counts			
Edges	−0.07		0.05
2-stars	0.23		0.05
3-stars	0.76		0.09
Triangles	0.74		0.24
Alternating k -stars	−0.07		0.05
Alternating k -triangles	−0.07		0.05
Alternating k -two-paths	−0.07		0.06
Degree distribution			
Standard deviation	2.15		0.76
Skew	1.56		0.40
Clustering			
Global	2.40		1.33
Local	1.59		0.85

degree distribution, we included a Markov 2-star parameter in addition to the previous higher order parameters. The parameter estimates are provided in the second panel of Table 3, with goodness of fit statistics in the right column of the third panel. This model does a good job of reproducing the standard deviation and skewness of the degree distribution, and improves clustering as well. The 2-star parameter estimate is positive and significant¹⁶ and the estimate for the alternating- k -star parameter has compensated by becoming more negative than was the case in the previous model. The k -triangle parameter is still significant and positive. Although the k -star parameter is not significant, its much larger negative value suggests that, allowing for the positive 2-star, there is more segmentation in this network than might have been surmised from the previous model. The positive 2-star reinforces the “popularity” of high degree nodes (there are many 2-stars in a hub), thereby producing a somewhat longer tail on the degree distribution than was the case for the previous model.¹⁷ So the impression is of a rather centralized triangulated structure, with that

¹⁶ The small value of the 2-star parameter should not be misunderstood. This reflects the fact that the number of 2-stars in a graph may be rather large and is not an indication of effect size.

¹⁷ Subsequent simulation confirmed that the positive 2-star parameter has the effect of producing a distribution of graphs with greater propensity for high degree hubs, with somewhat flatter and more dispersed degree distributions.

centralization in part due to some high degree nodes (the 2-star effect), but nevertheless with the possibility of some segmentation (the negative – but not significant – alternating k -star effect).¹⁸

6.2. Comparisons of maximum likelihood and pseudo-likelihood estimates

The UCINET data enable a comparison of the Monte Carlo maximum likelihood (ML) estimates with the approximate pseudo-likelihood (PL) estimation procedure. There has been a limited comparison of the two estimators in previous studies (see Wasserman and Robins, 2005, for a summary).

We compared PL and ML estimates in the models for UCINET data by scaling the difference between the PL estimate and ML estimate as a ratio of the ML standard error. In examining the PL estimates for convergent models, woflk had PL estimates which were clearly inadequate, with an extremely large and negative reciprocity estimate and a massive PL standard error. This result was clearly occasioned by the absence of mutual ties in the network (see footnote 13), and accordingly we exclude woflk from the following results. The results for the scaled differences between the two estimates for all other convergent models are presented in Fig. 8.

From Fig. 8, we see that most PL parameter estimates were within two standard errors of their Monte Carlo maximum likelihood counterparts, and many were within one standard error. In summary, 11 of the 12 convergent Markov models had PL estimates for all parameters within the 95% confidence intervals of the ML estimates. This is perhaps not surprising because, if Markov models converge, then dependence among the network variables is unlikely to be strong, and in these circumstances it is expected that PL estimates are more likely to be consistent with maximum likelihood results. Pseudo-likelihood is somewhat less reliable for higher order models where about one-quarter had PL estimates that would be considered extreme compared to the ML estimates. Considering the ratio of PL to ML standard errors in these data sets, the median value was 0.91 with a range from 0.29 to 4.87. So the PL standard errors are slightly smaller than ML standard errors on average, but they can be smaller or larger by a factor of three to four.

Sometimes it is very evident that PL estimates may be questionable, for instance, when they are extremely large. When this is not the case, problems with convergence for Markov models appear to occur when the star parameters are close to zero (suggesting the edge/triangle model of Fig. 1), or where there were no negative star effects. Whenever a Markov model did not converge and a triadic parameter was significant according to the PL results, then an alternating k -triangle parameter was significant in the resulting convergent higher order model. This result reinforces the attention paid to higher order transitivity by Snijders et al. (2006).

Overall, these results suggest to us that the PL estimates do give some information about the data, but their values cannot be relied on. Probably the best that can be said is that PL estimates that suggest “significance” according to PL standard errors may be indicative of the effects needed to model the data. The results reinforce our claim that ML estimation procedures should be the method of choice.

¹⁸ Models that include both 2- and 3-star Markov effects along with the higher order parameters also converge. Both 2- and 3-star parameter estimates are positive and the alternating k -star parameter estimate is negative, reinforcing the interpretation given here. But none of the effects are individually significant, not surprisingly because the alternating k -star parameter is a weighted sum of Markov star parameters, so the inclusion of sufficiently many Markov parameters merely explains similar variance in the data.

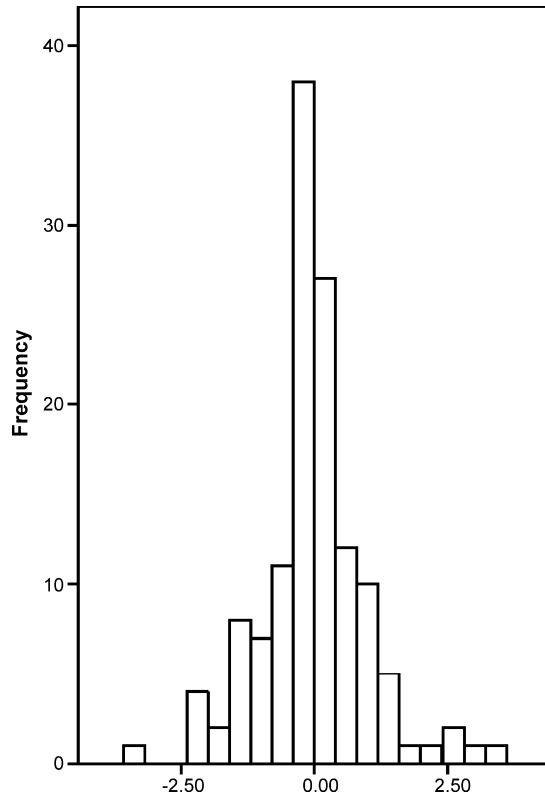


Fig. 8. Scaled deviation of pseudo-likelihood estimates.

7. Conclusions

In this article, we have reviewed the new specifications for exponential random graph models proposed by [Snijders et al. \(2006\)](#). We demonstrated by empirical example how these may be more effective than Markov random graph models. This is not to suggest that Markov random graph models are necessarily inappropriate. The point, rather, is that the new specifications offer ways forward in the quite frequent cases when Markov graph models fail because of near-degeneracy. We have illustrated how Markov parameters may still be used effectively alongside the new specifications, and indeed Markov dependence is still an important component in the new models. We do not claim that these models will be effective for all networks. But the new models do illustrate how different parameterizations may be developed to address particular modeling purposes.

So there is further work to be done. The focus in this article (and to a lesser extent in [Snijders et al., 2006](#)) has been on non-directed graphs. The recommendations by Snijders et al. for new parameterizations of models for directed graphs have already shown their value in achieving convergence in the case of models for directed graphs in the UCINET data sets. But these recommendations can be elaborated further in interesting ways. This is a matter for ongoing work.

Moreover, there are still many gaps in a full representation of network structure. This is evidenced by some of our goodness of fit examples, where a convergent model still does not adequately capture all of the features of the observed network. In addition, the contribution of

the k -two-path statistic needs to be better understood. We have also not discussed the inclusion of nodal attributes in these models. It is not difficult to incorporate node-level social selection effects into the higher order models (Snijders et al., 2006, provide examples).

The new parameters involve higher order constructs than Markov graph configurations. They implement the partial conditional dependence approach of Pattison and Robins (2002) which emphasizes longer-range connectivity rather than direct adjacency. That such a higher order approach is often necessary for successful modeling of empirical networks may tell us something of the emergent properties of certain social systems. It seems clear that sometimes Markov configurations do not simply accumulate homogeneously across a network; rather, they agglomerate heterogeneously within regions of the network. This fact emphasizes the connectedness and overlapping nature of the basic structures (such as triads) that we use to analyze observed networks. Through such approaches, then, we may begin to appreciate the interpolation from local to global structures, and may seek further understanding of the emergence of higher order structural features.

Acknowledgements

This work was supported by grants from the Australian Research Council. We are grateful for comments from the social network groups at the University of Groningen and the University of Melbourne, and for the helpful suggestions of an anonymous reviewer.

References

- Boer, P., Huisman, M., Snijders, T.A.B., Wichers, L.H.Y., Zeggelink, E., 2006. StOCNET: an open software system for the advanced analysis of social networks. Version 1.7. ICS/Science Plus, Groningen.
- Borgatti, S., Everett, M., Freeman, L., 1999. UCINET 5 for Windows: Software for Social Network Analysis. Analytic Technologies, Harvard, MA.
- Butts, C.T., 2002. Memory Structures for Relational Data in **R**: Classes and Interfaces Working Paper. Department of Sociology, UC-Irvine.
- Butts, C.T., 2005. The Network Package. Manual. Department of Sociology, UC-Irvine (URL: <http://erzuli.ss.uci.edu/R.stuff>).
- Frank, O., 1991. Statistical analysis of change in networks. *Statistica Neerlandica* 45, 283–293.
- Frank, O., Strauss, D., 1986. Markov graphs. *Journal of the American Statistical Association* 81, 832–842.
- Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo maximum likelihood calculations (with discussion). *Journal of the Royal Statistical Society, Series C* 54, 657–699.
- Goodreau, S., 2007. Applying advances in exponential random graph (p^*) models to a large social network. *Social Networks* 29, 231–248.
- Handcock, M.S., 2003a. Statistical models for social networks: degeneracy and inference. In: Breiger, R., Carley, K., Pattison, P. (Eds.), *Dynamic Social Network Modeling and Analysis*. National Academies Press, Washington, DC, pp. 229–240.
- Handcock, M.S., 2003b. Assessing degeneracy in statistical models of social networks. Center for Statistics and the Social Sciences. Working Paper No. 39.
- Handcock, M., Hunter, D., Butts, C., Goodreau, S., Morris, M., 2004. Statnet: An *R* Package for the Statistical Modeling of Social Networks. Manual, University of Washington (URL: <http://www.csde.washington.edu/statnet>).
- Hunter, D., 2007. Curved exponential family models for social networks. *Social Networks* 29, 216–230.
- Hunter, D., Handcock, M., 2006. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15, 565–583.
- Park, J., Newman, M., 2004. Solution of the 2-star model of a network. *Physical Review E* 70, 066146.
- Pattison, P.E., Robins, G.L., 2002. Neighbourhood-based models for social networks. *Sociological Methodology* 32, 301–337.

- Pattison, P.E., Wasserman, S., 1999. Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology* 52, 169–194.
- Robins, G.L., Pattison, P.E., Kalish, Y., Lusher, D., 2007. An introduction to exponential random graph (p^*) models for social networks. *Social Networks* 29, 173–191.
- Robins, G.L., Pattison, P.E., Wasserman, S., 1999. Logit models and logistic regressions for social networks, III. Valued relations. *Psychometrika* 64, 371–394.
- Robins, G.L., Pattison, P.E., Woolcock, J., 2005. Social networks and small worlds. *American Journal of Sociology* 110, 894–936.
- Snijders, T.A.B., 2002. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure* 3, 2.
- Snijders, T.A.B., van Duijn, M.A.J., 2002. Conditional maximum likelihood estimation under various specifications of exponential random graph models. In: Jan Hagberg (Ed.), *Contributions to Social Network Analysis, Information Theory, and Other Topics: A Festschrift in Honour of Ove Frank*. Department of Statistics, University of Stockholm, pp. 117–134.
- Snijders, T.A.B., Pattison, P., Robins, G.L., Handcock, M., 2006. New specifications for exponential random graph models. *Sociological Methodology*.
- Snijders, T.A.B., Steglich, C.E.G., Schweinberger, M., Huisman, M., 2005. *Manual of SIENA Version 2.1*. ICS, University of Groningen, Groningen.
- Strauss, D., 1986. On a general class of models for interaction. *SIAM Review* 28, 513–527.
- Strauss, D., Ikeda, M., 1990. Pseudo-likelihood estimation for social networks. *Journal of the American Statistical Association* 85, 204–212.
- Wang, P., Robins, G., Pattison, P., 2006. *Pnet: A Program for the Simulation and Estimation of Exponential Random Graph Models*. University of Melbourne.
- Wasserman, S., Pattison, P.E., 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika* 61, 401–425.
- Wasserman, S., Robins, G.L., 2005. An introduction to random graphs, dependence graphs, and p^* . In: Carrington, P., Scott, J., Wasserman, S. (Eds.), *Models and Methods in Social Network Analysis*. Cambridge University Press, New York, pp. 148–161.